DOCUMENT RESUME

ED 103 446 TM 004 277

AUTHOR	Mann, Peter B., Ed.
TITLE	Assessment in Schools, Colleges and Society.
INSTITUTION	Educational Testing Service, Princeton, N.J. Center for Statewide Educational Assessment.
PUB DATE	Nov 73
NOTE	48p.
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Colleges; *Educational Assessment; Educational Testing; *Evaluation; *Evaluation Needs; Higher Education; *Mass Media; Occupational Choice; Program Evaluation; Testing
IDENTIFIERS	*Needs Assessment; Social Implications

ABSTRACT

Testing, assessment, and evaluation, all based on measurement, are assuming increasingly important roles in American society. They are becoming a more vital force in the world of work, in education, and in the processes of determining the effectiveness of social institutions. Educational Testing Service (ETS) convened a conference for education editors and writers on "Assessment in Schools, Colleges and Society" because newspapers, magazines, and journals, after all, are the primary conduits for information flowing from specialists to the various groups affected by their work. But communication with the general public, educators, and business and government leaders is essential for specialists in measurement. The success of their work depends to an extraordinary degree on the understanding of both those whose abilities and achievements are measured and those who use the results. Some topics of discussion were evaluating educational programs, forecasting assessment needs in higher education, assessment for career decisions, and roles and implications of assessment for society. (Author/DEP)

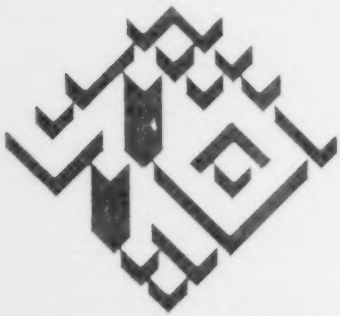
ED103446

BEST COPY AVAILABLE

ASSESSMENT

IN

SCHOOLS, COLLEGES & SOCIETY



report of a conference for education editors and writers

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
DOROTHY URBAN
ETS
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

TM 004 277

Edited by Peter B. Mann

November 11-12, 1973
Henry Chauncey Conference Center
Educational Testing Service
Princeton, New Jersey 08540

Educational Testing Service is an Equal Opportunity Employer.
Copyright © 1974 by Educational Testing Service. All rights reserved.

BEST COPY AVAILABLE

CONTENTS

Introduction 5

Measurement and Its Uses 7
William W. Turnbull

Where Are We Going? 12
Samuel Ball

Evaluating Educational Programs 17
Samuel Messick

Statesmen, Scientists, and Soothsayers:
Forecasting Assessment Needs in Higher Education 22
K. Patricia Cross

Assessment for Career Decisions and Roles 28
Winton H. Manning

Implications of Assessment for Society 34
Melvin M. Tumin

A Look at Education Reporting 39
Fred M. Hechinger

Some Points of Interest 42

Roster of Participants 46

Introduction

Testing, assessment, and evaluation—all based on measurement—are assuming increasingly important roles in American society. They are becoming a more vital force in the world of work, in education, and in the processes of determining the effectiveness of social institutions. Educational Testing Service is an organization devoted to the responsible development and use of measurement technology and techniques. In this time of change, then, it was appropriate for ETS to convene a conference for education editors and writers on *Assessment in Schools, Colleges, and Society*.

Newspapers, magazines, and journals, after all, are the primary conduits for information flowing from specialists to the various groups affected by their work. That is true of specialists of any kind, but communication with the general public, educators, and business and government leaders is essential for specialists in measurement. The success of their work depends to an extraordinary degree on the understanding of both those whose abilities and achievements are measured and those who use the results. And the stakes are often high: college admission or employment for the individual, gain or loss of public confidence for the institution.

The 1973 conference was enlightening. ETS staff members provided the journalists with up-to-date information on the status of, and prospects for, measurement; the journalists gave staff members fresh insights into the layman's view of their field. The conference was also enjoyable. Its setting was the new ETS Henry Chauncey Conference Center in Princeton, and the atmosphere was pleasantly informal. For staff members, there was the added satisfaction that the conference was one of the several events marking the 25th anniversary of ETS—and that the focus was not on the successes of the past but the challenges of the future.

This report is an effort to record the essence of the conference, to provide not only easy recall for the participants but information for those who could not attend. There is no attempt here to capture all that was said. The papers presented by the seven main speakers, reproduced in full, are followed by a brief summary of the major concerns, *Some Points of Interest*, that emerged from the discussions.

Unfortunately, I was unable to attend the conference because of a throat infection. My paper, *Measurement and Its Uses*, was given—and improved—by Scarvia Anderson, an ETS vice president and director of the Atlanta Office and co-chairman of the conference with Jane D. Wirsig, executive director of Information Services and Publications.

When she finished reading my paper, Mrs. Anderson noted that it contained a definition of *measurement* and said this word and three others—*testing*, *assessment*, and *evaluation*—have specialized meanings and would be used often during the conference. She proceeded to offer quick definitions, and at the next session they were amplified by another ETS vice president, Samuel Messick. The distinctions between these related terms are certainly as important to the reader as they were to the conference participants. So here is an amalgam of the definitions:

Measurement is systematic observation of behavior under specified circumstances, leading to results that are expressed in meaningful terms, usually quantitative, and that can be reproduced at different times and places. Measurements can be taken in

many ways—by tests, questionnaires, interviews, observations, ratings, instruments that record physiological responses, and so on.

Testing is one kind of measurement in which specific instruments (often of the paper-and-pencil variety) are used to gauge an individual's ability, achievement, interests, personality traits, or other characteristics. Usually, a test focuses on a single variable, but several variables can be measured simultaneously by administering a battery of tests or an omnibus test involving subscores. In the aggregate, test results can be used to characterize groups, such as classrooms or schools.

Assessment is a more comprehensive process. Frequently, several tests are involved as well as questionnaires, interviews, and observations, in order to obtain a broad picture of a person, program, or situation.

Evaluation adds to assessment the element of judgment. After the assessment results are in, the evaluators apply standards of worth or acceptability to them—for example, do the results indicate that this program is effective in accomplishing its purpose? Tests and other measures, while they may provide the basis for evaluation, are not in themselves evaluations.

It is hoped that the reader, armed with this knowledge, will proceed with relative ease through these papers. If so, the reward should be a broader view and a deeper understanding of measurement in its multiple manifestations.

William W. Turnbull
President, ETS

Measurement and Its Uses

By William W. Turnbull

In any discussion of testing or measurement, or assessment, one of the problems is that those terms conjure up quite different things to different people.

It is as if someone said let's discuss *means of transportation*, and one person thought about space craft, another about airplanes, another about trains, and so on through bicycles and pogo sticks. There are probably some generalizations to be made about all these forms of transportation for a particular purpose.

At a high level of generality, the conversation may go along quite well. Suppose one man thinks *train* while another thinks *pogo stick*. They may agree that it beats walking, that fatigue sets in after a while, that the metal ones are much better than the wooden ones, or even that it's great to hop on one—but fierce and futile arguments may break out when one party says they should be nationalized.

I'd like you to visualize several different kinds of situations. Last Saturday in Omaha, Nebraska, more than 200 high school students took the Scholastic Aptitude Test. By definition, that's a test. Last week, in our Infant Laboratory, Michael Lewis patiently showed a four-month-old child a series of colored slides while recording the infant's pulse, respiration, and EEG. Although the child didn't know it, her degree of attentiveness (or boredom) was being measured and related to the kind of picture being shown.

In California last spring, more than 100 colleges made use of something called the Institutional Goals Inventory—a questionnaire designed to elicit the aspirations of students, faculty, administrators, and trustees for the same institution. Here, the object of scrutiny is an institution rather than a person. Just the same, the process is broadly one of measurement.

And so it goes. Measurement encompasses the very young and the very old; traditional study, as in a classroom test of spelling, and nontraditional, as in assessment of the educational value of a young person's Peace Corps experience. It refers to attributes both intellectual and temperamental. It takes different forms and conditions, depending on whether it's for guidance, evaluation, amusement, self-understanding, selection, placement, or credit—and on whether an individual is to be measured or a group of people, an institution, a school system, a course of study. I always liked Scarvia Anderson's paper on how to measure the drawing power of a museum display. Its title is "Noseprints on the Glass."

You may be getting the idea that I include any observation at all under the measurement rubric. Not quite. Measurement must be systematic enough to possess several attributes. The circumstances under which the behavior or events occurred must be specified, the characteristics under observation need to be identified, the results must be expressed in meaningful terms (whether quantitative or not), and all the procedures must be described and followed with enough precision so that you or someone else could repeat the whole process and come up with reasonably similar results.

Much of the research on measurement itself has had to do with how to achieve those characteristics using a wide array of techniques in a great variety of situations. The academic disciplines surrounding measurement have grown up around just such questions.

Through thousands of studies, the answers have begun to come in. The results point clearly to the measurability of a very wide spectrum of human abilities, interests, values, and attainments, although the measurement techniques that work best vary depending on what you want to measure. (Of course, no sensible person would have thought otherwise.)

The kinds of tests that have been most prominent in popular discussions over the past couple of decades have been intelligence tests and the scholastic aptitude and achievement tests used for college entrance.

The IQ has worked a great deal of mischief in education and society. The reasons are several but stem mainly from the unfortunate and quite erroneous interpretations that have come to be associated with the term. The IQ is assumed to be an innate characteristic, which is certainly false in considerable measure, although nobody knows how much. It is believed to be unchanging through the person's life, which it is not. It is looked upon as a global, unitary trait, which certainly is untrue; questions on IQ tests usually tap a number of different skills and abilities. Its score points have an aura of exactitude they do not deserve: like all other measures in the human domain, they are approximations, afflicted by random error. Two different kinds of IQ tests may give markedly different results, quite systematically, over and above the random error.

For those and other reasons, you may not be too surprised to learn that IQs and IQ tests are held in rather low regard in ETS quarters. In fact, ETS neither makes nor gives IQ tests, nor do we report IQ-equivalent scores. If none of you wish to mention IQs, you will not be disappointing anyone here. We are, of course, glad to discuss them, but you need have no fear of offending if your attitude is critical.

Measuring Particular Abilities

Measures of particular abilities and attainments, on the other hand, are somewhat less laden with built-in errors of interpretation, although certainly they are not immune to them.

Back in the 1920s, Carl Brigham devised something he called a *scholastic aptitude test*—a rather general label, but one that he considered much more specific than the label *intelligence test*. His purpose was solely to measure the extent to which a student possessed the abilities needed for successful pursuit of college-level studies. There was no implication that these abilities were inborn or immutable. Moreover, the composition of the test was experimental, even though it was guided by Dr. Brigham's hunches about what college students ought to be able to do.

He had some good hunches and some bad ones. Among the latter was the guess that good students would be better able to read people's emotions from their expressions than would poor students and so he included a test section in which the problem was to pick a word that described the emotion being expressed by a face shown in a line drawing. When he followed up to see whether or not there was

in fact a correlation, he found there was none. The ability to read expressions apparently had nothing much to do with scholastic success.

Among his good guesses was the idea that ability to reason with material expressed in words—or verbal skill—would be important in college study. It turned out that verbal skill was probably the single most dependable index of readiness to pursue a university program successfully. After a few years of further experimentation and follow-up research, he found that mathematical skills were similarly useful as an index of probable success in certain kinds of college courses.

Brigham's work was done largely in the 1920s and 1930s. Since that time, we have explored many varieties of measures for various purposes, but it has proved exceedingly difficult to find skills that have greater general utility than verbal and mathematical ability—if your purpose is to predict scholastic success as measured by college grades.

The statement I just made needs to be examined and tested in several different contexts, and this has in fact been the subject of a good deal of research. Let me pursue the logic for a few more moments.

First, it seems to hold true at all score levels. People sometimes say that if a student has an SAT score of 600, or some other magic number, he has all it takes—and other qualities, like motivation, will determine his success. I have no doubt that motivation is indeed exceedingly important, but it appears to operate at all score levels rather than to determine success only at the high end of the scale. A student who scores 600 is more likely to do well than one who scores 500. A 700 student is a better bet than a 600. An 800 is a better bet than a 700. The correlation runs right up the score scale.

Second, the relationship seems to hold up even when the group is heterogeneous as to such basic characteristics as sex, race, and economic circumstance. That is, by and large, a rich white student and a poor black student who make the same SAT scores have an equal probability of succeeding in the same college program. This conclusion is contrary to popular belief, and so I think it is important to put it on the table.

Now let me go on quickly to cite some inferences that should not be drawn from the preeminence of verbal and quantitative abilities in predicting college success. First, it should not be inferred that SAT scores are perfectly precise. In fact, they are not: it is most unlikely that an individual tested twice would receive the same score both times. Second, what I said about the likelihood of success in college refers to performance averaged over a group, not to individual performance. Any single person may do much better or much worse than others with similar scores. A third incorrect inference is that SAT scores measure all the characteristics that are basic to college success. That certainly is not the case.

A fourth unjustified conclusion is that SAT scores are perfectly fair to everyone, since any given score indicates the same probability of success for members of quite different groups. This means that the test relates closely to the requirements for success in college as it is presently organized and conducted. If, in college study, the dice are now loaded against some students and in favor of others, the loading is

probably the same in the SAT. As college study undergoes changes, it may well be that the measures used to predict success in college—measures that have worked well—will need to be revised or replaced with other kinds of measures.

I have spent this much time on the SAT not because it has overwhelming educational importance but because it has been the subject of a great deal of misinformation and misunderstanding. Over the past 20 years or so, the testing of individual students for the purpose of selection by an institution has commanded the greatest amount of public attention and discussion. This has resulted in part from the fact that the educational establishment was underbuilt, so there was a shortage of places in relation to the burgeoning demand for further education. Now the situation is radically different. Supply has caught up with demand. Although there are still a few schools and colleges that are turning students away, their number has dwindled rapidly. Recruitment has replaced selection as the preoccupation of most institutions.

Helping Students Make Choices

Under these circumstances, the kinds of measures that are needed are different, too. Increasingly, the need is for information that will help students make the choices that confront them—information about themselves and about the institutions they might want to attend. For colleges, the new need is for information to help them place students appropriately and to grant academic credit for learning accomplished outside the framework of regular courses. How much credit should a student be given if he or she has picked up the basics of accounting by working in a CPA's office but without taking a college course? What is the academic value of a year in the Peace Corps?

If all learning fell naturally into academic disciplines and subdivided neatly into courses, we would be well along the path to meeting this growing measurement need. The College-Level Examination Program (CLEP), pioneered by ETS and sponsored by the College Entrance Examination Board, represents a major accomplishment: the translation of informal learning into equivalent academic credits. For someone entering a standard degree program, CLEP clearly provides a valuable means of obtaining credit for applicable experience.

Amid the current concern about career education, it is interesting to note that the relationship of education and work moves in two directions, and measurement is becoming an increasingly important tool in both. I've already pointed out the rising need for measuring the educational value of work experience. That's one direction. The other is measuring the on-the-job value of education or training.

How well prepared is the law school graduate to practice law? How competent is the auto mechanic—whether trained at a community college or a vocational school—to repair cars? Certainly, measuring competence for professional certification and occupational licensing is not new, but it is being extended to new fields. Auto mechanics is one example. Others range from law enforcement to hospital

financial management to real estate. ETS is "into" these fields plus many others. One of our latest assignments, by way of illustration, is to develop means of assessing the proficiency of golf pros for the Professional Golf Association.

This increased emphasis on measuring the results of formal education and training brings to mind several related points. As I mentioned earlier, testing as a basis for admission to secondary schools and college is no longer growing. But it is far from finished. Some institutions will remain selective by choice so long as they remain in sufficient demand to survive. They are not yet—and may never be—threatened by higher education's transformation into a buyers' market. Additionally, selection remains a vital testing function at the graduate and professional level, especially in those fields where demand is still greater than supply. Law school and medical school are the readiest examples.

While colleges are seeking valid measuring devices for purposes other than selection—for guidance, placement, and awarding credit—these needs are also emerging earlier in the chain of formal education. Means of gauging individual learning progress are now considered important not only at the secondary and elementary levels, but in preschools and kindergartens. In fact, increasing attention is being focused on the assessment of preschool youngsters.

Aside from assessing the progress of students at all levels of education, there is a strong trend toward the use of testing to evaluate the effectiveness of schools, school systems, colleges, and programs such as *Sesame Street* and *The Electric Company*. This trend is clearly a response to popular pressure, both for the improvement of public education and for the increased accountability of the educational establishment to the taxpayers.

Without lingering too long on the assessment of institutional effectiveness as distinct from the assessment of individual student progress, I should cite the interrelatedness of the two kinds of effort. For the individual student, test results can be helpful—if they are used to guide advancement through the learning maze, tailor programs to needs and interests, and alert teachers to learning gaps. For the school or college, test results can provide an index of strengths and weaknesses in the institutional endeavor. If this index is used to improve the enterprise, the beneficiaries will be the students of the future. My point is that measurement of both types aims ultimately at improving education for the individual student, either today or tomorrow.

Where Are We Going?

By Samuel Ball

Let me tell you first of all what I am supposed to be talking about: *Educational Goals, Objectives, and Tests.*

I had originally entitled this speech "Hans Christian Anderson, A Visit to the Steel Works, and *Moby Dick*: Where Are We Going, Where Do We Think We Are Going, and How Will We Know When We Get There?" But I was told that was too long. So I shortened it to "H. C. Anderson and M. Dick," but I was told that was too cryptic. So we finally settled on "Where Are We Going?" That's a clear question–but I am not sure the answer is clear.

And now for the story of Hans Christian Anderson, though I doubt whether it is true. One day Hans was walking along the street when he met the king. "Hello, where are you going?" said the king in perfect Danish. "I don't know," said Hans vaguely. The king grew angry because kings like to have straight answers so they can lower the boom if they have to. "Don't give me that nonsense," said the king. "Where are you going?" "But I really don't know," replied Hans. Enraged, the king ordered H. C. Anderson clapped into prison for contempt of court, as it were. The following week the king visited Hans, still in prison. "Now are you willing to tell me where you were going?" And Hans replied, "I told the truth, for I did *not* know where I was going. How was I to know I was going to gaol?"

In 1955, I taught sixth grade at Clemton Park Public School, having graduated from teaching third an l fourth grades for the three previous years. I had to teach my sixth-graders about Australian secondary industries, so one of the things I did was to take them on a trip to the steel works at Wollongong in the Illawarra district south of Sydney. We left early in the morning and got home pretty late, and we spent some days beforehand in preparation and afterwards on follow-up activities. A few weeks later, I had my annual visit by the district inspector of schools. "Why," he asked, "did you spend all that time on the visit to the steel works?"

"Because I wanted the class to know how steel is made."

"Aha," said he with a malevolent inspectorial glint in his eye. "Don't you realize that you could have *told* them how steel is made in much less time than taking them on this trip?"

"Yes, but they also learned how to get along with each other, and the need to thank their hosts at the steel works, and the importance of planning trips...."

And then my inspector pointed out that they were not the reasons I had taken the children to the steel works. "You are shifting your ground," he pointed out with some logic.

Moby Dick is one of those classics that is frequently ruined in high school. Paul Diederich of ETS tells the story of the time he was visiting a high school class where the teacher was very proud of his difficult examinations, which he saw as being indicative of his high standards. The class had been studying *Moby Dick*, and Paul was intrigued at the questions on the class examination. "Name and describe the various species of whales encountered by Captain Ahab." "Name and give an account of the ships encountered by Captain Ahab and do so in the order they were encountered." Well, it could have been worse–on another exam was the question, "What was the name of Polly's cat in *Tom Sawyer*?" (And last year my daughter

came home with her social studies test on which was one chestnut, "When did Columbus discover America?" and one atrocity, "Which troops in the Revolutionary War were not brave?" The answer, I discovered, was "the British.")

Goals and Objectives

Let me return later to Hans Christian Anderson, a visit to the steel works, and *Moby Dick*, and let me now talk about goals and objectives. I like the convention of using "goal" to denote the long-term destination we seek to arrive at toward the end of an educational journey. Goals are important because, if we don't have the long-term destination in mind, the educational processes we design will most probably become a spaghetti-like, chaotic set of worms.

Of course, it is impossible not to have some goals in mind, but it *is* easy to keep them implicit. Once we try to make them explicit, the sins of our thinking are laid bare. It has been fashionable since John Dewey's *Democracy and Education* was published over 55 years ago to talk about our pluralistic society. But one problem with a pluralistic society is that it enables us to espouse many competing and contradictory goals all at the same time. And as the Mormons discovered, there are problems when you espouse too much at one time. Take, for example, the days before *Sesame Street* went on the air, when the goals of the show were still being considered. There seemed to be general consensus that the show should teach preschoolers sharing and cooperation. But a black psychiatrist raised the issue—should black children learn to be cooperative and to share in general? What about when they were being put upon? Should they cooperate then? What if they had been systematically excluded? Should they share at most on a selective basis—qualified sharing?

Or consider, in real life, the goals of doing one's duty and respect for legitimate democratic authority. Remember recently General Haig telling Mr. Ruckelshaus, "Your Commander-in-Chief orders you to do so." And remember the response—"No!"—and the subsequent firing? Certainly the majority of America wants to educate children to respect authority, but it also wants its children to be moral and have high principles. In schools, teachers want their students to learn what they're being taught, but they also want their students to be questioning and creative. Some of us want our girls to grow up to be good wives and mothers, while others of us want our girls to grow up to be good human beings (and let them subsequently choose the specific roles they want to assume). In any case, there is conflict among our goals, partly because we live in a pluralistic society and different groups have different desires, and partly because within ourselves we have not thought through adequately what it is we want to do.

I will use "objective" to denote the short-term destination—each of the milestones we must pass in order to reach the goal. Objectives are important because without them we have no useful way of knowing whether we are on track. When we look at objectives, we see the conflicts among the goals in even clearer focus. We

want our children to be cooperative, generous, and independent (one set of goals), and we also want them to have good school spirit (another kind of goal). We set up an objective to have a first-rate high school or college football team that will beat the opposition, deny the opposition the chance to score; and all this in the name of generating "school spirit." But we ignore the effects on the other goals, generosity and independence, and we don't even consider the question: If the goal is cooperation, with whom should we learn to cooperate—our friends on our football team or the strangers from the other school's football team?

Of course, it is easy to point out these inconsistencies between and among goals and objectives. We want good health for all our students but put a vast share of the money for our physical education program into the needs of a few students so we can have good school teams. We want our students to learn to be good citizens, but we give them a school so organized that they get little opportunity to practice the skills of a good citizen. Actually, I think our objectives are much nearer to the reality we want in our society than the nice-sounding goals. Do we really want our citizens thinking critically, behaving independently, working creatively? What would this do to our government and our factories? (Maybe this should be thought of as a cynical aside.)

Which brings me to the point of considering the tests we present to our children. These, for most students and teachers, are the real expressions of our goals and objectives. And here there is even worse cause for disquiet. We test what we find easy to test. We find it easiest to test for achievement of factual knowledge and relatively easy to test for the achievement of understanding of that factual knowledge. But even here there is an inherent problem that we must face—the problem of validity. Are the achievement tests we use actually measuring the knowledge and skills we think they are? Just because a test is labeled "mathematics," is it measuring achievement in the mathematics we are teaching in the schools? Are the reading tests whose results are so widely published (and so outrageously fudged) in fact testing reading achievement? They may be—but they may also be assessing anxiety level, ability to follow complicated directions, vocabulary level, and perceptual skills. Have you ever looked critically at a reading test? Are the directions appropriately written for a third-grader? Is the test timed? What kinds of questions are asked? Are they related to the instructional objectives and educational goals? How many items are provided on the one page? We simply don't look hard enough at the test items and the tests from which we make our generalizations about school achievement. Thus, we think it is relatively easy to test for cognitive and academic achievement, and it is. But it is not as easy to do these things well.

Attitudes and Feelings

We find it even more difficult to assess attitudes and feelings, and we often ignore completely the need to assess non-test behavior. We often work on the tarnished

principle that if students know what to do, they will want to do it. But we know that teaching students so that they know and comprehend what is in the Bill of Rights will not necessarily mean they will support the bill's precepts when it comes to specific issues. And we know that knowing the story of the good Samaritan does not ensure that those so knowledgeable will behave like the good Samaritan. Why do we teach children civics? So they can show they know what is in the lessons? Or so they will feel and behave in certain ways? As your friendly Blue Cross and Blue Shield association says: "There's more to good health than paying the doctor's bills." So, too, there's more to good education than passing tests of knowledge and comprehension. We have means of assessing other domains, but we tend to neglect the use of these means.

Just as I was finished writing this speech, I got a letter from my nephew in Australia and it was so appropriate I thought I would introduce it here (in part, of course).

> Dear Uncle Sam,
>
> I'm enjoying school but I don't like the school work.
>
> Once a week we have civics lessons. Miss Smith tells us about some bloke that was real good. I like the stories but I can't remember everything about them, and I get the questions wrong at the weekly test. I can never remember what country Father Damien came from or what year it was when Albert Schweitzer was born. It really is a big problem. Miss Smith told me I was a bad boy and I'd never make a good citizen unless I pulled my socks up. Well, it's true I hate civics, but I do want to be a good citizen. I hate civics more since she kept me in to learn stuff about Florence Nightingale yesterday and that made me real mad because I'd promised to help the Matron at the Wagga Wagga Methodist Home for Orphans. I'm going to make some blocks for the little kids. The man in the hardware is letting me have some paint, and I've organized some of the other kids at school to help me do the painting. I've got an idea about making other toys with three ply and fretsaws after we've made the blocks. I sure hate civics, though. I have shop, too. Last week we had to draw a capital I in a book and then do it a second time on coloured paper and cut it out. I ruined my job because in the drawing book I put my base line one inch up from the bottom instead of two inches. So I got an F for shop.*

So where are we going? Hans Christian Anderson pointed out truly that he did not know where he was going. I doubt whether, despite all the work carried out on goals statements (and they are very numerous), we understand where we are going

*This letter is based on "The Poor Scholar's Soliloquy" by Stephen M. Corey, published in *Childhood Education*, Vol. 20 (January, 1944). Pp. 219-220.

in education either. Part of the reason is that there are external pressures that will put us off course anyway, rather as the king moved Hans off course. But another part of the reason is that we have been rather soft-headed. We have not confronted the moral, ethical, psychological, social, and political issues that arise when we look at the conflicts among legitimate goals in a pluralistic society. We glory in comprehensive, melting-pot high schools with common cores in the curriculum for all students and then point out how good it is to individualize instruction.

And where many think we are going (our goals) may not be where we are in fact going because our objectives simply may not be attuned to our goals. As a famous attorney general of the United States—John Mitchell—once said, "Judge us not by what we say, but judge us by what we do." By the same token, look at what is happening to find out where we are really going in the schools. (Do not look too long at where we say we are going.) The fact is that the reasons for the trip to the steel works for class 6B in 1955 could have been best seen by looking at what was happening in the classroom (the daily objectives in operation). I still think that inspector wasn't exactly fair.

And finally there's *Moby Dick* and my Australian nephew and the problem of tests and assessments. And how will we know when we have reached our goals and objectives? While I admit that *Moby Dick* and my Australian nephew provide examples of the worst kind, they also suggest better ways. We can assess an appreciation of *Moby Dick* better than our rigorous teacher, and my nephew has presented, unwittingly, some ideas on indicators we could use to test the effectiveness of our education.

As I see it, I have raised a number of important educational questions and problems about goals, objectives, and tests. I have not answered these questions nor have I solved the problems. To do so at this point would hardly be fair. After all, we have the rest of this day, five more speakers and eight more discussants to go. And even if we cannot answer and solve them all in this time and with these resources, maybe there is some good in having raised them anyway. It might point the way to where we should be going.

Evaluating Educational Programs

By Samuel Messick

There are two major reasons why we might want to evaluate educational programs. One is for purposes of program improvement and system renewal: to gain increased understanding of processes both producing and impeding effectiveness so that rational changes might be made to improve program functioning. The other is for purposes of accountability: to gain information about the relative effectiveness of programs so that rational administrative decisions might be made in allocating resources.

Not coincidentally, there are also two methodological approaches to evaluation extant in the technology of the field. One is called *formative evaluation*, or evaluation for program development and betterment. The other is called *summative evaluation*, or overall appraisal of a final product. Given two approaches and two purposes and the obvious congruence between formative evaluation as an approach and program improvement as a purpose, it is easy to fall into the position of accepting summative evaluation as the natural approach for the other major purpose of evaluation, accountability.

In this paper I will argue against this tendency to tie accountability to the overall appraisal of program outcomes, especially as typically implemented in what might be called the *economic* model of program evaluation, which focuses upon gains or input-output differences relative to cost. It is true that this approach does enable us to ascertain the degree to which a program works and to cull out those programs that do not work or are not worth the cost. What better basis could we have for determining accountability? But the time perspective is wrong for constructive accountability. This summative approach is oriented toward the past. It tells us that a program *was* relatively cost effective and hence is probably worth continuing—provided, of course, that circumstances do not change. In this framework, accountability is primarily negative—educators may be held responsible for what happened, but they are not given much help in determining why it happened or what to do to rectify matters.

What is needed is an approach to accountability that is oriented toward the future, that goes beyond an overall assessment of the size of effects to an investigation of the processes that produce the effects. Only an understanding of these processes will provide a rational basis for changing programs to improve effectiveness and to accommodate to changing conditions.

What is needed is an accountability system that views decision making as problem solving, that provides an empirical groundwork for diagnosing inadequate educational progress and for prescribing plans for corrective action. In this framework, accountability is basically positive—educators are held responsible for taking corrective action when goals are not achieved. Such an approach has been suggested in the ETS design for an accountability system for the New York City school system (1).

From the present perspective, educational evaluation should involve research on both process and outcome in order to accrue information relevant for both program improvement and administrative decision making. At first glance this may seem like a blurring of the useful distinction between formative and summative evaluation, and to some extent it is—but there really is no fundamental conflict involved. The

terms *formative* and *summative* refer to different roles of evaluation, not to different forms of evaluation, and data from a particular evaluation study can and usually should play both roles. Indeed, what is needed is a system of continuous evaluation entailing recurrent formative and summative elements to provide an ongoing basis of current understanding from which to make rational and adaptive changes. Such a system, as we shall see, must include a comprehensive measurement and information program that goes beyond a consideration of educational treatments per se to an assessment of the broader social context within which those treatments operate.

The Need for A System Perspective

Inherent in the concept of accountability is the notion that the individual who is accountable is a causal agent. But educational effects are multiply determined and sometimes overdetermined. Given multiple causes, there are also likely to be multiple agents responsible for these effects, whether desirable or undesirable. No equitable system of accountability can ignore these complexities and interdependencies. In all fairness, you can only hold a person accountable for what he does or fails to do, not for what someone else does or fails to do. This phenomenon of multiple causation of educational outcomes introduces considerable complication into our efforts to understand program functioning, for it requires us to unravel the sources of causal influence in assessing responsibility for outcomes. This problem highlights the need for a system perspective in educational evaluation.

At this point, then, it may be helpful to make explicit some properties of the educational system that are usually either taken for granted or overlooked. To begin with, the educational enterprise is, indeed, a system in the technical sense of a complex set of elements that functions as a whole by virtue of the interdependence of its parts. There are many parties to the system, and many parts and processes. The system is composed of differentiated but overlapping subsystems that embrace the student, the family, the community, and various peer groups, as well as the school, the teachers, and the programs. Since the concept of system implies a functioning whole whose various elements and subsystems are interdependent, it follows that a change in one part of the system may produce unanticipated and possible adverse consequences in another part of the system. Thus, it is not enough to evaluate an educational program solely in terms of its expected effects – we must also assess a wide range of possible effects. Indeed, the side effects of an educational program are frequently more important than the intended effects.

And the situation is even more complicated than this, for education or schooling is not only a system but a culture. The existing features and regularities are determined and maintained by roles and expectations, attitudes and values, tradition and history (2). Under these circumstances, an understanding of program effects requires not only knowledge about person and program variables but also about social and institutional variables.

One implication of this recognition of complex system interdependencies is that educational evaluation should routinely adopt a multivariate interactional strategy: it should employ multiple measures in each domain or subsystem and methods of analysis sensitive to interacting influences (3, 4). Within this framework, research questions become elaborated from straightforward comparisons, like "Is this new program more effective than the old one?" to something more complicated, like "Do these programs or program components interact with personality and cognitive characteristics of the students or with factors in their educational histories or family backgrounds to produce differential effects upon growth?" Note that such an elaboration is important, for if trait-by-treatment interactions occur (that is, if different people respond in different ways to different kinds of programs), or if background factors turn out to moderate program effects, then the typical simple comparison of average gains for different groups will very likely be misleading.

In contrast to the economic model of evaluation mentioned earlier, this systems approach has been characterized as a *medical* model, in that it emphasizes side effects as well as intended effects and an understanding of causal processes as well as the gauging of treatment outcomes (5).

Some might object at this point that this particular view of educational evaluation is overly elaborate and complicated, that with all this emphasis on process and interacting influences it runs the risk of subverting the main aim of evaluation, which is an appraisal of outcome or product. This latter stance has come to be called in some circles the "butter-wrapping model" of evaluation (6): if you have taught someone how to wrap butter, the main thing you want to know is how many pounds of butter he can wrap. No matter if different processes of butter-wrapping are employed by different butter-wrappers (one may use his right thumb, another his left forefinger). What matters is how many pounds of butter he can wrap after the training experience. But suppose one very effective butter-wrapper has gained his speed and efficiency because of a particular stylistic quirk; namely, wetting his thumb with his tongue as he picks up the waxed paper. Would you be very happy with that particular style of butter-wrapping, even though it produces more wrapped butter than any other? At least in some cases, then, we must take into account factors of style and process in order to evaluate the very desirability of the outcome.

Continuous Evaluation: A Basis for Adaptive Action

Ideally, what is needed is a systematic basis for understanding not only the change processes operating in programs per se, but the moderating factors and interactive forces operating in the broader context of the educational system. We need methodologies for objectively determining and describing basic characteristics of system functioning and for evaluating both the changes in those characteristics and the efficacy of the change process in relation to intended and unintended outcomes of the system and of intervention programs.

One possibility would be to introduce into school systems a continuous program of information collection and analysis to uncover existing regularities and processes and to monitor their perturbations over time. Such an information and assessment program would have to be comprehensive enough to assess a wide range of student outcomes representing intended and unintended consequences of the educational system, as well as characteristics of teachers, students, administrators, programs, and settings that might interact with each other to produce differential results. It would also be particularly important to include some provision for observing and documenting those most critical of all regularities occurring between teacher and student in the classroom. Given the cultural basis of much system functioning, it would be vital to incorporate procedures for assessing the attitudes and values, roles and relationships, perceptions and expectations, and aspirations and goals of the various parties to the system and to attempt to clarify dimensions of consensus and contention permeating the social matrix.

With such a multiplicity of components, it is obvious that the program would have to include an analytical methodology capable of multivariate and longitudinal comparisons to provide mechanisms for disentangling the threads of multiple and interacting influences. Ultimately such an assessment effort should attempt to go beyond the description of system functioning as it exists to the development of causal models relating prior conditions and processes to outcomes, so that the potential consequences of alternative educational programs could be systematically anticipated.

A new instructional program or intervention project could be introduced into an educational system being monitored in this fashion with prior warning about existing social pressures and with periodic feedback enabling adjustments and improvements in program procedures. The impact of the intervention could then be systematically evaluated, in straightforward if not routine manner, in terms of changes in measures of system characteristics, especially student outcomes.

In the accountability system proposed for the New York City schools (1), for example:

> The first step is to gather data on as many school and community characteristics as possible. Two criteria are used for selecting these variables: (a) they must be factors the school system can do something about; (b) there must be reason to expect that changing them will affect student learning. (These are called) *process variables* because they describe processes that affect instruction and learning. The process variables suggested for initial study fall into five categories: characteristics of the school's staff, characteristics of the school's students, characteristics of the school's programs, characteristics of the facilities, and characteristics of the community. To illustrate: a staff characteristic might be educational experience; a student characteristic, reading habits; a facilities characteristic, age of school building; a program characteristic, total expenditure on programs; a community-district characteristic, percent of teachers with more than three years' experience.

A comprehensive educational system information program of the type envisioned here represents an ideal that would require considerable effort to realize. It offers the advantages of a wide range of information bearing on administrative and instructional decision making, program development and evaluation, and accountability, and in the long run could serve as a vehicle for continuing research on the functioning of educational systems. As a continuing program, it would also provide an early warning system signaling the appearance of new types of students and new or changing conditions, thereby permitting timely modifications in programs to accommodate to new inputs and circumstances. Its major disadvantages are the time and resources required to develop, implement, and maintain it. Other potential difficulties derive from the fact that such a comprehensive assessment program would itself become an important part of the system it is meant to monitor and hence might influence or modify that system in unforeseen ways. Thus, even techniques introduced to monitor system functioning and to facilitate informed decision making may produce more pervasive effects and should themselves be examined in terms of system interdependencies, in particular with respect to the psychology, economics, politics, and ethics of information collection and use.

This concept of continuous evaluative research is intended to provide, at the system and program level, a rational basis for adaptive action. In addition, however, much of the information involved in this ongoing evaluative function, such as measures of pupil progress, could serve both evaluative and instructional purposes. Thus, if systematically implemented with feedback at the teacher and classroom level, this conception could help make evaluation an integral part of the educational process itself.

References

1. McDonald, F., & Forehand, G. "A Design for Accountability in Education." *New York University Education Quarterly*, Vol. 4, 1973. Pp. 7-16.

2. Sarason, S. B. *The Culture of the School and the Problem of Change*. Boston: Allyn & Bacon, 1971.

3. Messick, S. "The Criterion Problem in the Evaluation of Instruction: Assessing Possible, Not Just Intended Outcomes." In M. C. Wittrock & D. E. Wiley (eds.), *The Evaluation of Instruction: Issues and Problems*. New York: Holt, Rinehart, & Winston, 1970. Pp. 183-202.

4. Shulman, L. S. "Reconstruction of Educational Research." *Review of Educational Research*, Vol. 15, 1970. Pp. 371-395.

5. Messick, S. "Evaluation of Educational Programs as Research on Educational Process." In F. F. Korten, S. W. Cook, & J. I. Lacey (eds.), *Psychology and the Problems of Society*. Washington: American Psychological Association, 1970. Pp. 215-220.

6. Wittrock, M. C., & Wiley, D. E. (eds.). *The Evaluation of Instruction: Issues and Problems*. New York: Holt, Rinehart, & Winston, 1970.

Statesmen, Scientists and Soothsayers: Forecasting Assessment Needs in Higher Education

By K. Patricia Cross

The exhilaration and frustration of the assessment business is due in part to the fact that evaluators exist on the cutting edge of change. Agencies such as ETS are shaped by, and in turn help to shape, the directions of educational innovation. Because it takes a long time to develop methods and tools for assessment, we must often predict educational directions before they are fully articulated by the educational community. On the other hand, we can't be so visionary that we develop materials educators are not ready to use. The tension between the push and pull of responsiveness and leadership is constant, and at the practical level, timing is critical. In the past, ETS has made errors of being both too late and too early. But we have also had our share of the sort of good timing that serves the educational community well. Let me give one example of a test development program that was started far enough ahead of the demand so that it was well developed by the time educators were ready to use it.

Ten years ago, in 1963, ETS started the development of what has become the College-Level Examination Program (CLEP), designed to help adults achieve credit for learning outside the classroom. Not so incidentally, that was also the year when Harold Wilson made a bid for the intellectual vote in England by proposing the Open University. Open University, as you may know, is designed to provide educational opportunity for thousands of Britishers who never had the opportunity for education beyond the secondary school. It concentrates on delivering education to people who, because of life circumstances, cannot study full-time on a college campus. The mission of Open University is educational opportunity, and the attractiveness of that concept was fully apparent but not universally articulated in 1963.

The existence of such a successful model as Open University, however, helped to stimulate worldwide interest in external study and lifelong learning. Assessment of out-of-class learning is a crucial accompaniment of the external degree, and the timing of CLEP helped immeasurably to spread the acceptance of credit-by-examination in this country. Development of the program, remember, had been quietly under way for five or six years before nontraditional study and the external degree became a red-hot innovation. Because of good timing, the CLEP exams were ready to meet rising demands for credit-by-examination and to stimulate even greater interest in proficiency testing by their very availability. Today, proficiency on the CLEP exams is accepted by two-thirds of the colleges in the country as an alternate to classroom experience.

Who knows what the ingredients for successful timing really are? Actually, the CLEP tests were probably a little ahead of their time. They were devised because they seemed a good and educationally sound idea, not because there was a readily visible market. But we could also cite some examples of good educationally sound ideas that never caught on. ETS was ready to provide tools of assessment that would help education, but there was no adequate market for them.

Testing agencies frequently serve a fluctuating role between stimulating innovation and responding to it. The availability of assessment tools makes possible some reforms, but the best tools in the world won't spark a reform that educators are not

ready to implement. Those who worry about testing agencies dictating educational directions have never tried to keep a program afloat in a nonprofit corporation until educators wanted to use it.

I think I can make the case that timing for ETS requires the presence of scientists, statesmen, and soothsayers in the organization. Professionals at ETS must play the role of educational statesmen and reformers by recognizing educationally sound ideas and helping to influence the development of the tools that would make possible their implementation. Our role as scientists and educational researchers is to understand the implications of data that extend the frontiers of knowledge about the educational process. And our role as soothsayers is to recognize, even before they articulate it, what colleges and students are going to want. My personal focus at ETS has been on serving students. As researcher, I need to know what they want. As statesman, I need to know what they need. And as soothsayer, I need to know what educators are going to give them.

Determining What Students Want

Let us talk first about the role of scientists or researchers at ETS in determining what students want in education. Dick Peterson, a research colleague in the Berkeley Office of ETS, conducted some research last year at the request of the California State Legislature (1). He administered the Institutional Goals Inventory (IGI) to faculty, students, administrators, trustees, and citizens of the local communities of 116 private and public colleges in California. Each constituent group was to rate their college as it *is* and as it *should be* on 20 different institutional goals. The results indicated that students in the various types of colleges were notably dissatisfied with two aspects of their present education. The extent of their dissatisfaction was shown by the discrepancy between the *is* and *should be* ratings of their college.

Highly able students at eight campuses of the academically selective University of California felt that too much attention was given to the acquisition of knowledge in the academic disciplines. They said that, as a goal, academic development of students presently ranks third in the list of university priorities whereas they would demote it to 14th place in the list of 20 institutional goals. Their fellow students in the community colleges agreed that even in their less academically rigorous institutions, the emphasis on academic content was greatly overdone. They would demote it from its present first priority to ninth. When students across the broad spectrum of academic accomplishment reject so emphatically the major activity of colleges and universities, we have some thinking to do. Are students failing to perceive the importance of our efforts, or have we lost our perspective about the mission of higher education? Further analysis of Peterson's data sheds some light on the question.

It is quite clear that students are not expressing anti-intellectual attitudes. They are strongly supportive of activities that place more emphasis on the development

of the appreciations and skills of learning. They recognize as important such goals as these: instilling a commitment to lifelong learning, creating the desire for self-directed learning, and developing in students the skills of analysis and synthesis. The message that seems to be conveyed by this research is that we need to give more attention to the development of intellectual skills and attitudes and perhaps less to the simple purveying of information. Thus, as scientists we know something about the desires of students and their frustration with their present education.

As statesmen we can only agree that education for content mastery is a very short-range objective. With the speed of change today, content knowledge is outdated by the time a student moves from freshman to senior—even assuming that professors are up-to-date on the newest developments in their field. Instilling a love of learning and a flexible response to change, rather than content mastery, should be the fundamental task of a college education. Difficult as the project is, the educational leadership must begin to tackle that job, and we need the tools of assessment in order to do it.

Finally, as soothsayers, we need to make some guesses about the educational programs that colleges are going to provide. One way of injecting some science into our prediction of what educational institutions are going to do is to look at the priorities of the power structure—the faculty and administration—of colleges. Referring again to Peterson's data, it is clear that almost everyone—faculty, students, administrators, and lay community people—would give the development of broad intellectual skills and interests higher priority than they now have. Thus, it is a very good guess that if we, as evaluators, could develop convincing measures of these attributes, there would be a ready market for them, and we would be contributing to improved educational practice.

The future of tests directed toward assessing content mastery in the academic disciplines is not so clear. While students are obviously dissatisfied with the present overemphasis on content mastery, faculty, administrators, and governing board members of the universities of the California system fail to see any huge discrepancy between what is and what should be in this area. Thus, a valid short-range prediction would seem to be that traditional academic achievement tests will enjoy continued use in the universities because the people who determine educational programs there value content mastery and emphasis on knowledge in the academic disciplines.

The situation in the community colleges, however, is quite different. Community college faculty and administrators are almost as disillusioned about the goal of academic development as students are. Their top priority goals are the creation of a feeling of community on campus, the vocational preparation of students, and the personal development of students. The academic development of students is given only a 12th-place ranking by both faculty and administrators in community colleges today. As soothsayer then, it would seem to be a reasonable prediction that community colleges will lead in de-emphasizing assessment of academic content mastery in favor of acceptance of a greater range of tools of assessment.

Attending to Personal Development

Let us turn now to another source of student dissatisfaction with present educational programs. Students in all kinds of colleges are expressing a marked dissatisfaction with the present state of affairs regarding attention to the personal development of students. Personal development is defined by the IGI as helping students to do the following things: identify and pursue the accomplishment of their personal goals, develop a sense of self-worth and self-confidence, achieve deeper levels of self-understanding, and develop open and trusting relationships with others. University students claim that the university gives this goal 18th priority whereas it should rank fourth. Community college students think attention to student development is even more important; they would rank it second in the list of priorities. Community college students get a great deal of support from their faculty and administrators for greater attention to noncognitive growth as a responsibility of the college, but university students get very little. University of California faculty would grudgingly move emphasis on personal development from its present 14th place only to 10th place, and administrators are only slightly more interested. The power structure of the university thus stands in marked contrast to the students—and to lay members of the community, who agree with students about the importance of personal development as a function of a university education.

As soothsayer, I predict that evaluators could be quite helpful to community colleges by developing some good measures of noncognitive growth. While I see no immediate market in the universities, I might go so far as to say that noncognitive measures are a good bet for a long-range forecast even in the universities because the present dissatisfaction of students, supported by the general public, is likely to have an impact on the bastions of university traditionalism. If the measures prove successful in the community colleges, there is every probability that they will be desired by four-year institutions.

So far I have talked mostly about the changing needs of the present educational community. My analysis has been applied to students currently in the system. Now I would like to take a look at students presently outside the system, but waiting to get in. For if higher education has any growth potential at all, it is in those segments of the population previously unserved by higher education. The present leveling off of college enrollments is not a temporary phenomenon. The actual number of five-year-olds dropped 15 percent between 1960 and 1970. These are the potential college students of 1982. The number of births dropped 12 percent between 1970 and 1972. These are the potential college freshmen of 1990, and the nation's birth rate is now slightly below zero population growth (2). In other words, the college students of the next 17 years have already been born—or, of more significance to my argument, not born—and there is no expansion possible except to pick up some of the 40 percent of high school graduates who do not presently enter college or to serve the growing adult market of higher education.

Looking at the New Students

An analysis of students new to higher education shows these groups on the rise: 1) Students from the second and third academic quartiles of the high school graduating classes. These students are, of course, flooding into open admissions colleges. As a matter of fact, in the past two years, 85 percent of all the *increase* in the number of first-time students entered community colleges. 2) The proportion of ethnic minorities and women is on the rise, whereas the proportion of white males entering college immediately following high school graduation is on the decline. It dropped from 63 percent in 1968 to 53 percent in 1972 (3). 3) The proportion of adults returning to college is on the rise, and if the nontraditional studies movement continues its precipitous pace, this will be the most rapidly increasing segment in all higher education. So let us look at what these new learners want—because in a sense they are in a good position to get what they want. It is now a buyer's market, and colleges are likely to provide the programs that will attract the new learners.

First, let's look at the group that I, in my own research and writing, have called *new students* to higher education. These are primarily low academic achievers admitted to higher education by open admissions. Despite all the publicity given to ethnic minorities as new students, it is really the low-achieving white sons and daughters of blue-collar workers that are the chief beneficiaries of the movement of full equality of educational opportunity as reflected in the community colleges. As a matter of little known fact, about two-thirds of all the students in remedial education in community colleges are white, not black. These students are not in college for the love of learning or for the liberating influence of higher education; they are entering community college because they see it as a pathway to better jobs. They are not eager for liberal arts education except as it will give them respectable credentials, and they are heavily career-oriented.

The career orientation of these groups of low academic achievers is matched by ethnic minorities and women. They, too, are primarily interested in access to higher education as an entree to the mainstream of American economic society.

Finally, it is quite apparent from research collected by ETS on the interests of out-of-school adults between the ages of 18 and 60 that adults want an education that is useful. They will be attracted to college if the education provided helps them to develop the competencies that will be useful in their daily lives—to get a better job, to invest money wisely, to cook and garden and raise children, to repair things around the house, and to enjoy their leisure time.

The groups of students that represent the growth potential for postsecondary education, then, are interested in learning how to do things—as opposed, perhaps, to how to think about things. This suggests, to the scientists among us, a need for greater attention to what has been called competency-based education. If we can develop the tools that will tell students and employers what kinds of tasks people can perform, we will be meeting a need that is already emerging as a particular result of the increasing proportion of new students in colleges.

As soothsayers, we may be a little late in marshaling our resources to meet this need, which has already been articulated by the educational community. But as statesmen, we recognize the educational soundness of education and evaluation to help people improve their performance toward a specified goal—as opposed to education and evaluation that is basically competitive and tells people where they stand with respect to others. Normative testing is an essential tool for selective purposes, and it served selective admissions in the meritocratic years of the 1950s and '60s very well. But supply in higher education has now caught up with demand, and in our new egalitarian decades, educators are shifting attention from the task of selection to the job of education. Thus, assessment is moving out of the admissions office and into the classroom. In serving the new students to higher education we have an example of the dual role of evaluation—to be responsive and to demonstrate leadership. Ideally, we will be responsive to the new interests in the educative process, and we will offer the tools and services that will help to lead education to new levels of accomplishment.

References

1. Peterson, R. E. *Goals for California Higher Education: A Survey of 116 Academic Communities.* Prepared for the Joint Committee on the Master Plan for Higher Education, California State Legislature. Berkeley: Educational Testing Service, 1973.
2. Glenny, L. A. "The '60s in Reverse." *The Research Reporter,* Vol. 8, No. 3, 1973. Pp. 1-4.
3. Young, A. "The High School Class of 1972: More at Work, Fewer in School." *Monthly Labor Review,* June 1973.

Assessment for Career Decisions and Roles

By Winton H. Manning

A turn of a phrase that is strangely jarring to the reader sometimes captures the rapidity of change in our way of life that paragraphs or pages cannot do half so well. So it is that at least once or twice a month an obituary in the *New York Times* for some fallen business or professional leader will say, and I quote an example: "Mr. Smith prepared at St. Paul's for Yale, Class of 1924, and completed his education at Yale Law School in 1927, whereupon he entered the practice of law in New York City."

Few concepts are embedded as deeply in our culture as the notion that work is the object of life. Consequently education has been regarded as preparation for work, or even in the not too distant past for a career befitting one's "station in life." Hence, as Robert Hutchins said, "(Education) has been thought of as a children's disease; having had it once, you need not, in fact you cannot, have it again. This attitude has been reinforced by the organization of educational systems into stages: as each stage is reached, the one that is left behind is 'finished.' And, if an education is regarded as instrumental—to a job, marriage, a degree—its purpose has been fulfilled when its object is attained (1)."

Hutchins goes on to quote Margaret Mead on this same point:

> "...our educational institutions are places where we keep children for a shorter or longer period...Once they have left, we regard them as in some sense finished, neither capable nor in need of further education, for we still believe that education should come all in one piece, or rather in a series of connected pieces, each presented as a whole. ...Thus we avoid facing the most vivid truth of the new age: *no one will live all his life in the world into which he was born, and no one will die in the world in which he worked in his maturity* (1)."

What does this mean for educational institutions and programs? For Hutchins and others, it means radical reformulation of our ideas about when, where, and with whom education takes place. In short, if we are to live satisfying lives in the decades ahead, we must somehow create a "learning society."

Educational measurement, as a servant of the educational process, is necessarily a victim of the same dogmas of the past that dominate popular conceptions of our educational system. If education is seen as a "series of connected stages, through which one successively passes," it is no accident that examinations have served primarily as bridges from one stage to another, probably culminating for the recently deceased Mr. Smith in his bar examination, after which he could consider himself finished with both education and tests. How different such a life-style is from that of many persons now in their maturity, to say nothing of young people.

Just two weeks ago, for example, my wife received notification from the State of New Jersey that she is at long last licensed as a clinical psychologist. Stretching back from that official letter, which gives her legal sanction to enter into private practice as a professional psychologist, is a 20-year trail of examination papers—including college admissions tests, graduate admissions tests, Ph.D. qualifying examinations, doctoral orals, licensing examinations and so forth—but, more

importantly, all this was interspersed with babies, three cross-country moves, and all the vicissitudes of working one's way through an educational system that has never really seriously evaluated its procedures from the standpoint of a professional woman in our culture. Her first response to the letter was not, "At last! I'll never have to take another examination as long as I live!" Rather, her reaction was, "Now I can get busy learning more about some of the things I'm interested in!" A basic assumption of this paper is that we are going to see fundamental changes in postsecondary education over the next 25 years, which will make access to postsecondary education open, continuous, and more interrelated with employment and leisure than ever before – in short, a much different world than you and I have grown up in.

What are some of the ways in which educational measurement can serve to support this concept of lifelong learning – of a "learning society," to use Hutchins' term?

Appraisal of Relevant Learning

Clearly, there is an enormous job to be done in developing sound procedures concerned with the assessment of learning that goes on in experiential settings – that is, with competencies that are acquired independently by learners of any age, through work or other life experience. The boundaries separating formal educational institutions and the wider world of learning through experience are now breaking down, and the need is growing rapidly for new approaches to the appraisal of relevant, demonstrable learning that has been acquired in nontraditional study. For example, we are now attempting to organize a consortium of about 40 agencies and institutions who want to work cooperatively together, and with ETS, in developing measurement services appropriate to these new needs. Even five years ago such an effort would have been unthinkable.

If I am correct in asserting that the educational system is resolutely moving outward in the direction of what Hutchins first called a "university without walls," it is also changing in another dimension to which I referred earlier, namely, education is becoming a lifelong experience. Again, Margaret Mead stated the case well when she said:

> Often, only a few months may elapse before something which previously was easily taken for granted must be unlearned or transformed to fit the new state of knowledge or practice. In this world, no one can complete an education (2).

The world of Mr. Smith, like himself, has been decently interred. In its place, we must design new educational arrangements for a new world of learning. That is true whether we speak of liberal learning necessary to function as a competent citizen, or of specialized knowledge necessary for a productive, satisfying career as an auto mechanic, a clinical psychologist, a dentist, or even a journalist.

In this respect, ETS has been actively working with a number of groups who recognize that the rapidity of change requires continuous access to education. One of these is the National Institute of Automotive Service Excellence, which sponsors the auto mechanics certification program. Others are the American College of Dentistry and the American Board of Professional Psychology, to name a few examples. In each instance we have worked with specialists in the field to develop examinations that are voluntarily being taken by practitioners as a means of self-appraisal. The examinations become, if you will, a framework or grid of knowledge and competencies against which the auto mechanic or the dentist can measure himself. The results of these examinations are useful in helping the individual direct his own continuing education or training, and they also can assist those concerned with devising programs of training or continuing education that meet the current needs of individuals more effectively.

In some instances, notably the auto mechanics program, the results of the examination also can be used as a vehicle through which the individual can communicate the level of competency he has acquired to the public at large. When Mr. Smith graduated from Yale, his name was inscribed on the rolls of the Class of 1924 and his degree was conferred with the full panoply of medieval splendor. The chevrons and certificates awarded to the auto mechanic by the National Institute constitute comparable recognition, as well they should in a society that dignifies the worth of the individual and esteems competence however acquired. Furthermore, it is worth noting that Mr. Smith was never faced with the necessity of demonstrating in 1934, to say nothing of 1964, that he was keeping up with the times in order to retain his degrees! Examination programs are, therefore, beginning to serve as the handmaidens of continuing education within a context of lifelong learning.

Education for Survival

There is, however, a more fundamental aspect of education for which the rapidity of change requires that we rearrange our views and restructure the assessment services that will be supportive of educational processes in a learning society. This is education for personal development, if not survival, in modern society.

One of the consequences of the American conception of education as preparation for work is our inability to distinguish between education and training. Thus the objective of the educational system is to yield *trained manpower* rather than to produce *educated people*. This is reflected in the development of vocational guidance services in this country, which has relied heavily on the techniques of aptitude testing and related assessments.

On the whole, vocational guidance testing proceeded on the assumption that there exists a large but finite number of jobs that can each be described in terms of the abilities, knowledge, and interests necessary to perform them. The task of measurement specialists is one, therefore, of sorting or classifying people by these

assessment techniques into the appropriate job-related hierarchies, and then "guiding" them in the appropriate direction through counseling. The technical problems can be formidable, but in some applications, such as the military services, very sophisticated systems have been developed along these lines. When I was a graduate student, this kind of model was implicit in every industrial psychology course I took.

Unfortunately, this model of education and of occupational assessment does not deal well with the dynamic character of jobs today. The facts suggest that in a large number of occupations the pattern of abilities, skills, and knowledge can shift continuously as a function of rapid, technologically induced changes in job demands. Furthermore, as James O'Toole points out in *Work in America*, the report of Elliot Richardson's Special Task Force, about three-quarters of the jobs in this country today require relatively little advanced training, and what is needed in this regard can be accomplished in two or three days of on-the-job training (3).

Consider what may be the most symbolic but humble task of mankind—the baking of a loaf of bread. Hutchins reports the existence of bakeries in Germany where Spanish women ride to and fro on bicycles, watching signal lights on fully automated bake ovens that carry on the baking process from the raw material to packaged loaf, perceived only by electronic sensors (1). Or take the case of the Lordstown plant, GM's automated showcase, which inspired the prophetic title of Emma Rothschild's book, *Paradise Lost: Decline of the Auto Industrial Age* (4). What are the ability demands of such jobs? More importantly, what are the needs for education in order to live in such a world?

"Work," my industrial psychology professor told me 20 years ago, "is a form of activity that has social approval and satisfies a real need of the individual to be active. To produce, to create, to gain respect, to acquire prestige, and incidentally to earn money—these are some of the reasons why people work." The lack of some of these things is, we are told, the reason why young workers struck the Lordstown plant in February 1972. It also demonstrates the folly of considering education as simply preparation for a job, and the necessity for considering education as something more than training.

To return to the point made earlier, the failure of the American educational system to distinguish between education and training produces too many people who are not only ill-prepared to cope with the world of work as we now see it taking shape, but ill-equipped to enjoy the benefits of increased leisure that technology is beginning to provide. How is this connected to the problem of developing assessment services that are responsive to the new work environment? The older model of sorting people into the "right" slots is defeated before it begins because of the tempo of change. At best it will increasingly come to apply only to upper portions of the job hierarchy that require complex professional and technical knowledge. In other words, we have on our hands a vocational guidance model that has been rendered obsolete by the march of events, and which can justify its survival only by becoming increasingly elitist in the clientele it seeks to serve.

Guidance in Career Decisions

At ETS, Martin Katz has been leading an effort to develop a new approach to vocational and occupational guidance–one which emphasizes the *processes of decision making*, and which has as its objectives increasing the competence of an individual to understand himself, improving his ability to acquire and use information intelligently, and enabling him to make informed decisions for himself. Known as SIGI–System of Interactive Guidance and Information–Katz's approach uses sophisticated computer technology within a framework of humane educational objectives. SIGI is concerned with education and not training, in the sense that education is, in the last analysis, concerned with encouraging the *habit of thinking*, the capacity to distinguish between the important and the unimportant, and the *development of critical standards* of thought and action. In Martin Katz's words:

> Our bias–our conviction–is that in education enlightened processes are intrinsically important. Therefore we bend our efforts to increase the student's understanding of the factors involved in choice (imperfect though our understanding may be) so that he can take responsibility for his own decision making, examine himself and explore his options in a systematic and comprehensive way, take purposeful action in testing hypotheses about himself in various situations, and exercise flexibility in devising alternative plans (5).

New approaches to assessment for career decisions and roles are in the making, and SIGI is a preeminent example. Career assessment must go hand in hand with changes in educational systems and in the world of work, all three of which are inextricably intertwined. Whether there will emerge out of this ferment a new and exciting period of profound change in the philosophy and methods of education, assessment, and work is hard to say at this time. But ETS is engaged in the effort to make assessment serve the individual, thus contributing to this emergence.

Perhaps, in an economy where work as we have known it will become increasingly less important, it will be possible to implement the sort of system of education envisaged by Dewey in 1916, when he wrote:

> Both practically and philosophically, the key to the present educational situation lies in a gradual reconstruction of school materials and methods so as to utilize various forms of occupations typifying social callings, and to bring out their intellectual and moral content (6).

Hutchins goes further to the point when he adds:

> (Man's) educational institutions have been designed largely to perpetuate existing values. Ultimate recognition of the facts of life may force reconsideration of those values and the redirection of education toward new ones. The

> first step is general understanding of the facts of life, of the new values that are now attainable, and of the possibilities and limitations of education in helping to achieve them (1).

We have a long way to go to reach the objectives set forth by Dewey and Hutchins. The task for the assessment community is to take firm hold of these new directions or, like Mr. Smith, look forward to a long, long sleep beneath the sod.

References

1. Hutchins, R. M. *The Learning Society*. New York: Praeger, 1968.
2. Mead, M. "Why Is Education Obsolescent?" In R. Gross (ed.), *The Teacher and the Taught*. New York: Dell, 1963. P. 271.
3. *Work in America: Report of a Special Task Force to the Secretary of Health, Education, and Welfare*. Cambridge, Mass.: MIT Press, 1973. Pp. 134-152.
4. Rothschild, E. *Paradise Lost: Decline of the Auto Industrial Age*. New York: Random House, 1973.
5. Katz, M. "Exploring Values for Career Decision Making: A Computer-Based System of Interactive Guidance and Information." In *Measurement for Self-Understanding and Personal Development, Proceedings of the 1973 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1974.
6. Dewey, J. *Democracy and Education*. New York: Macmillan, 1916.

Implications of Assessment for Society

By Melvin M. Tumin

The delightful vagueness of the global mandate I have been given—to assess "the implications of assessment for society"—both requires and permits me to interpret my task according to my own best lights. I must do so, however, with decent regard for what I judge to have been the intentions of the organizers of this conference. They know as well as I that I am neither psychometrician nor sociometrician. I take it, then, that I am asked to consider the activity called assessment, in its many manifestations and contexts, and suggest some obvious, and perhaps some not-so-obvious, significances of that activity for this culture and society. My questions thus become: How extensive is assessment as a cultural activity? What functions does it serve? What does it portend?

Taken in its broadest implications, the term—and the activity—*assessment* goes far beyond the boundaries of educational effort. If we focus for a moment on the diverse functions assessment serves in the field of education, and use those functions as a checklist for looking at other fields of cultural endeavor, it becomes apparent that every major institutional nexus of our society is subject to the same range of assessment efforts. Theatre, poetry, fiction, sculpture, painting, architecture, dance, and music—in short, all the arts. Fiscal policy, monetary policy, investment activities, fluctuations in the stock market, growth and fluctuations in the gross national product, bankruptcies and successes—in short, all the economic activities. Child rearing, family solidarity and rupture, sex roles, day care centers, courtship and marriage, abortions and contraception—in short, all the institutions connected with kinship and reproduction. Similar recitations can easily be made for politics, recreation, religion, health and welfare, science, stratification systems, race and ethnic relationships, and international affairs. The entire institutional framework of our society is deeply involved in assessment activities of all kinds.

What are these activities? They involve measuring what has transpired, predicting what might transpire, understanding why events have taken place as they have, determining the weights of different variables that are considered relevant, locating the significance of events for the society, considering future lines of policy and action—in short, every conceivable dimension of these institutional activities is a proper subject for some form of assessment. Add to these the enormous effluence of "how to" literature, including how to perform the traditionally most private activities, not excluding the traditionally most tabooed activities. "How to" is, after all, a fundamental kind of assessment. It involves an understanding of means-ends relationships, the resources needed to perform a task, the training required to develop skills, the attitudes requisite to full participation—the whole kit and kaboodle. Even Linda Lovelace's talents are available for "learning" in various "how to" manuals.

We can now see this activity of assessment, in all its ubiquity, on quite different levels of significance, though those levels are not necessarily hierarchically ordered relative to each other. Assessment represents, first, an attempt to order rationally and give structural coherence to human activities that otherwise would seem inchoate, and without form or pattern. Assessment is also obviously intended to provide humans with somewhat greater control over human affairs, and affairs of

non-human nature, such as floods, earthquakes, and the weather in general. Third, assessment functions to make it possible for humans to exercise some greater rationality in choice among numerous competing alternatives (e.g., college courses and movie offerings). Fourth, assessment, such as is involved in market research, serves to make it possible to increase the range of choices humans might have available among future goods and services. Assessment also provides an extraordinary playground for both common and elite participation in major cultural activities. The millions of Monday-morning quarterbacks who have expertly diagnosed the wins and losses in Saturday horse races and Sunday football games are an example in point. Every man his own expert—a crucial function in a society in which genuine expertise requires great and prolonged effort.

A Question of Control

If there is one common strain that runs through all these functions of assessment it is the theme of the management of diversity. But that is not enough. For it is crucial also, I think, to recognize that the very effort of assessment, the presumption of assessment, if you will, is the predication of human responsibility for human affairs, of human activity as the instrument of human destiny, of human capacity to direct human affairs, of final human control over human fate.

I deliberately repeat and stress the word *human* to contrast it with an earlier theme that was dominant just about 200 years ago, when no such presumptions and predications could have received much public reception. For until 200 years ago human fate was considered to be subject primarily to Divine guidance and rule, including, importantly, the variance in human conduct made possible by the Divine grant of free will. Under that system of beliefs, the actions we call assessment here, which involve humans peering at themselves and their behaviors, trying to know about, understand, and see significance in, and to speculate about, futures—all these manifestations of human self-reflexivity and human self-consciousness—could not be important in the life of the culture. They became legitimate and possible only when, at the end of the 18th century, it became increasingly appropriate for humans to consider themselves as legitimate objects of inquiry.

Seen in that light, *assessment may be understood as the modern functional equivalent of the traditional practice of religious adoration.* It is not that assessments today lead to paeans of praise for and affirmations of the nobility and glory of human responsibility for human destiny. To the contrary. The dominant tone of such assessments is often that of despair, even if the fundamental source of the energy behind the pursuit of ordering and understanding lies in optimism and hope. Rather, the analogy to religious adoration derives from the realization that adoration of God's handiwork was the ritualized manner in which the extraordinary diversity and mystery of human affairs were managed, namely, by ascribing their totality and significance to Divine intention. So, on the modern scene—that is, from 1800 until now—the dominant response to the new awareness of human

responsibility for human history, and of human triviality in the world of non-human affairs, has been one which involves increasingly refined efforts to specify – that is, to assess – the role of humans in their own destiny and, at the philosophical extremes, to try to create a language whereby the meaning of that role and of the act of assessment of the role can be formulated and uttered.

We see, then, that what is common to the apparently greatly different responses of religious adoration and secular assessment is one striking function served by both; namely, the management of diversity. This is not to say that religious affirmation and adoration and secular assessment serve only that function. It would be no problem at all for all of us here to specify another dozen functions which both adoration and assessment serve. But it is to say that at least a principal function of both is the management of diversity. Having said that much, one major difference in the two approaches becomes immediately apparent. Religious adoration serves that function by attribution of all diversity to an all-powerful Divine agency, leaving the diversity unanalyzed, unreduced and unordered. By contrast, secular assessment serves to make diversity manageable by reduction through classification, typologizing, factor analyzing, correlating, scaling; in short, by grouping diversities into manageable numbers of types, and measuring, comparing, and contrasting this reduced number of types so that informed choices can be made among them.

The introduction of the term *choose* or *choice* implies another fundamental difference between the era of adoration and that of assessment. It is the fact that a crucial feature that accompanies the transition from the Society of God to the Society of Man (with apologies to women) is the emphasis on the optionality of human existence and the possibility of change in the quality of that existence through human activity. Optionality unavoidably means preference. We can come to prefer and make such preferences meaningful only when we adopt the perspective that our fates are not foreordained or preformed or given by a more-than-human agency. Hence, we have the right to prefer and try to choose. Moreover, we have the capacity to make such choices, and we have the technology for informing such choices – namely, through comparative assessment of the varied offerings.

The theme of optionality and choice has a still deeper significance that must be specified, especially since it is so strongly antagonistic to the notion of Divine creation of human existence and activity. I refer to the implicit notion that in modern society humans see themselves not only as agents of their own present qualities and destinies, but *as active creators of their future qualities and destinies as well.* That is, we are not content with understanding the lawfulness of our present existences. Rather, we insist, too, on the power to create our individual and collective existences and states of being for the future as well.

That theme – of continuous, intelligent, rational self-creation of individual and social forms and functions – may be said to be central to educational assessment activities. For better or for worse, we seem quite fully caught up in the idea that we must understand human variability, not only to understand it, but to see what can be done about reducing it, or at least reducing its consequences when we consider

those consequences undesirable. So, too, we have visions of a desirable society; its work force; its human relationships; its goods and services; the quality of life for its citizens; the psychological states of those citizens, including their levels of anxiety, creativity, utilization of intelligence, and affective well-being. So we measure, test, compare, contrast: we assess, that is, the better to shape, fashion, and control our futures and those of others.

A New Antipathy

Having said that, we now can see more clearly why various assessment activities, such as those of the Educational Testing Service and the College Board, have come to be viewed with so much suspicion and disdain, however widely employed those various tests and evaluations continue to be. The crucial suspicion is that the assessments being performed will be utilized only, or primarily, to select out from the existing human stock those cohorts who measure up to certain specifications needed to preserve the going order of things. Put in the most invidious terms, the assessment of educational functioning is seen by those who decry it most as an instrument for the preservation of existing inequalities and the perpetuation of the privileges of existing elites and their offspring. The evaluations are seen as devoid of, or even antagonistic to, the idea that humans with proper effort can shape new futures for old populations and create new populations with new endowments, new capacities, new roles and functions.

It is a mark of the tempo of change in modern society that while educational measurement experts feel they have barely begun to understand how the human mind and its various faculties work, and why populations differ in these manifest faculties, some segments of the lay public condemn the instruments of assessment as though they were already full-blown instruments of class and racial bias and malice. These are the laymen who feel most adversely affected by the ways in which such evaluations have heretofore been utilized.

This new theme–this antipathy to assessment, especially among underprivileged sociocultural groups and their champions–is in some ways more like the earlier adoration of Divine omnipotence than the newer secular theme of comparison and contrast. For underlying the antipathy to assessment is the partly religious notion that all humans are equally entitled to an equal share of the good things of life, regardless of the differences in their measured talents and abilities. All humans, in short, are equal "children of God." Hence, all efforts to measure them, typologize them, categorize them, and assign them different shares of the good things of life are to be deplored. I have said that the underlying spirit is partly religious. But it is also partly Marxist and partly humanist, and, in the final analysis, it can be shown to be a direct outcome of certain philosophical and scientific works which also have led to the development of psychometrics and sociometrics. I am thinking particularly of the extraordinary way in which John Dewey, in *Human Nature and Conduct*, combines a respect for the fundamental equality of all human beings with

an equal respect for the importance of rational inquiry and measurement of causes and consequences in human affairs.

This leads me to assert that—all notions to the contrary notwithstanding—there is no inherent moral quality to assessment per se. It is a technological instrument. As such, it is value-neutral and non-directional. It can be and has been utilized for the most diverse purposes. But its availability and the state of its development do not determine its usages or direct human affairs. It is like all other technological products—like all other machines, if you will. Humans decide why they want machines, what kinds of machines they want, and how they will use them. If a machine seems to be operating *per manus Dei* instead of *per manus humanis*, the single act needed to bring the machine under control is simply to pull out the plug.

But that is only a temporary measure, and one born out of desperation. The fact that humans have great power to put assessment and other technological devices to the uses they themselves choose to serve suggests that some approach other than pulling out the plug would be desirable. If it is humans who finally determine how technology will be used, and it is, then it is humanly arranged social power that determines who will determine the uses of technology.

Assessment, in sum, has come to serve two enormously important functions, namely, to enable humans to manage diversity through reduction and ordering, and, simultaneously, to expand human existence and give it new and richer meanings by creating new ranges of options and choices, which can be made in more informed and rational ways because of the enlightenment that assessment provides. How fully each of these two major functions of assessment will be implemented is, finally, a human choice.

A Look at Education Reporting

By Fred M. Hechinger

A visitor from another planet would, had he happened to drop in on some recent Presidential press conferences, be forced to conclude that the American sector of Planet Earth would be devoid of problems if only the news media stopped creating them. Those anti-Nixon intellectuals (educators included) who react to the charges against the news media with anger, however, are not entirely honest, if they imply that the President has invented these defensive tactics. As any veteran education writer can attest, educators are past masters of the technique.

For example, during a recent meeting with New York City school administrators, a principal launched a spirited attack against *The New York Times* practice of publishing the reading test scores for all the system's schools. The scores, he said, lead the uninformed public to draw all kinds of erroneous conclusions about the quality of the schools. "Besides," he added, "the tests are no damned good."

Having been the villain who originally urged, authorized, and assigned the undercover operation that broke the secrecy that used to keep the scores hidden from public view, I had to rise to the defense of *The Times*. I suggested that publication of the information could hardly leave the public in greater ignorance than it had been without the information. I conceded that it is the newspaper's responsibility to accompany the scores with all the available explanatory and interpretive information. Such efforts, I suggested, had been made honestly and diligently, but if specific omissions could be brought to our attention, they ought to be corrected. Finally, I asked: "If the tests are, in fact, 'no damned good,' then why are they still being used?" (Except for the fact that we had repeatedly reported the experts' doubts about the tests in question, the principal might have triggered another kind of story which would h ve pleased him even less than the original one.)

The point I am trying to make is that educators, who tend to be politically very insistent on the people's right to know, are often not very consistent in applying this principle to their own activities. Just as Mr. Nixon has repeatedly complained that the media fail to report what is right with America, educators regularly charge that the media are wallowing, if I may coin a phrase, in what is wrong with education. Many who cheered the disclosure of the Pentagon Papers nevertheless denounced the disclosure of the reading scores—and that's only a random example.

I am afraid there is no answer that will satisfy those who ask: Why is bad news given so much more space than good news? Good news *is* given quite a bit of space, provided it is newsworthy and not merely routine. Few educators who are also commuters would hail a newspaper for commenting favorably on the fact that many trains *do* arrive on time.

The Role of the Media

Now for a few specific observations concerning what I consider to be the role and responsibility of the media in educational reporting.

My first answer to that question is simply reporting. The media ought to report what is happening (and, of course, what is not happening). They ought to tell what the establishment or the profession does—and where it succeeds or fails. They ought

to report what the non- or anti-establishment forces are doing, planning, criticizing, and proposing.

To accomplish the foremost task—reporting—the media ought to make available time and space in far more generous proportion to the rest of the news than is now the case. I am not saying this in a utopian, pie-in-the-sky manner, without full consideration of media economics; I am merely asking for a more fitting allocation of time and space *within the total scheme of coverage*. (If advice-to-the-lovelorn columns are an indispensable public service, why not advice-on-education columns, too? The case load of the education-lorn may, in fact, be greater than that of the lovelorn.)

A second vital function, closely related to reporting and crucial to successful reporting, is translating. Just as a science story is useless if it cannot be understood by the lay reader, so is an education report meaningless if it is comprehensible only to the profession. I won't bore you with a recitation of my 20-year struggle against professional jargon; I merely suggest that not enough members of the education profession will ever be taught comprehensible English to eliminate the need for education reporters who also are skilled translators. I realize that the translation often offends and even infuriates the professional "source" of a story; the cause for such displeasure is sometimes a faulty translation but more often the fact that the story, in translation, has lost its intended cover-up of the truth. That is precisely why good reporting and translation must go hand-in-hand.

The third crucial requirement is interpretation. It is the area in which the media are probably most vulnerable, partly because it requires space but more importantly because it requires qualified interpreters. (Let me come back to the latter requirement a little later.) What is needed is a constant effort to place education news in historical, national, and—even more important—international perspective. When I am told that the schools are the most oppressive institutions, I want to know in relation to what? To prisons? To corporations? To French schools? When I am informed that American schools have failed to close the inequality gap, I want to know how well or badly they have done on that score in comparison to British or German or Indian schools. I want to get some insight into the schools' historic mission—what was it? How well has it been carried out?

The final category in which the media have a crucial role to play is opinion. If I am somewhat hesitant to say much about this function, it is largely because that's how I earn my living now, and that, of course, means that things are just fine. Seriously, though, I think quite the opposite is true. The editorial pages—and that is where opinion ought to be confined—are more often the place where reactionary publishers and superannuated editors parade their nostalgic memories about the good old days when the schools turned out such paragons of virtue as themselves. The theme, more often than not, is antipermissiveness and the extolling of the old-fashioned virtues and the work ethic—a la Agnew, you know. Honesty is the best policy—for others.

Clearly, the "they don't make them like me anymore" line about education is a useless form of opinion. The less said about it the better—particularly the less said about it on editorial pages.

This leads me to an observation that covers all the categories—reporting, translating, interpreting, and opinion. Specialized personnel is essential, but unfortunately conspicuous by its absence. The trend is one of backsliding. The media are using fewer specialists in education today than they did a decade ago.

Let me underscore that by specialists I do not mean former teachers or graduates of teacher-training institutions or even people who have taken courses in educational journalism. What I am asking for is people who have devoted some years to education coverage, and who have done so by preference rather than in surrender to orders from above. A lively interest in education—free of any bonds of allegiance to any educational dogma and, even more important, to any educational organizations or groups—is all I ask for.

What I also ask for is that the media keep such home-trained specialists on the education beat unless they eventually see fit to promote them to desk editorships or other exalted positions. I suggest that not as a lobbyist for the specialists but as a means of getting education-conscious people into policy-making positions.

Unhappily, I must say that the course I am proposing is not currently being followed—not by the newspapers (including my own, which keeps shifting people into and out of the education news department before their experience has a chance to pay off), not by the newsmagazines, and least of all by the television networks, which, to the best of my knowledge, have no education specialists or even consultants. The networks used to do some occasional, and even some very fine, education documentaries, but these efforts appear to have been declining ever since the various revolutions—student revolt, urban confrontation, and curricular ferment—were called off for lack of money and interest.

That leads me to my final point—loss of interest. Education is currently in a deep slump. Its outside support forces appear to have scattered. The Nixon administration has undoubtedly viewed this development with satisfaction and is not about to rally the citizenry to the support of an enterprise that might cost money. The schools and colleges are shell-shocked by recent unrest and current recession. Few ideas appear to be coming out of the academic establishment. The period of growth and expansion is over.

The media cannot—in fact, must not—create action where there is no action. But the media are responsible for the coverage of education, and the city editors' argument that the public is no longer interested simply does not hold water.

Lack of progress and absence of planning and leadership are legitimate and important areas for coverage in education, just as lack of progress and absence of leadership are always considered legitimate and important areas for coverage in politics and government. And, of course, instances of action and leadership amid general depression and fatalism are that much more newsworthy.

Some Points of Interest

Following the speakers' presentations, a number of comments and questions were posed by the respondents and editors. These led to lively discussions in which several points of interest or concern were raised.

One speaker chalked a square on the board and rapidly divided it into quadrants. He labeled the columns *Educational Uses* and *Noneducational Uses*, and the rows *Formal Educational Experiences* and *Nonformal Educational Experiences*. Then he asked ETS staff members in the audience to rank the quadrants in terms of the current state of measurement science. In other words, he was asking: Is it possible to measure an individual's formal learning equally well for educational and noneducational purposes? Can informal learning be measured as precisely as formal learning?

One response: The greatest expertise is in measuring formal learning for educational uses, the least in measuring informal learning for noneducational uses; the relative expertise in the other two quadrants is debatable. "It's easier to agree on what makes a good law student," this staff member observed, "than on what makes a good lawyer."

After the session, another staff member dismissed the speaker's questions as an "academic exercise." While it certainly was that, it nevertheless stirred the interest of several editors and writers, and as the conference unfolded, the reason became obvious. There is widespread uncertainty about *the state of the science*. How precise are test scores? How accurate are they in predicting academic and occupational performance? How relevant are they to the purposes—admission to college, assessment of schools and colleges, hiring, professional certification and occupational licensing—for which they are used?

These questions were raised directly or indirectly, and surfaced throughout the conference. So did some of the answers. Measurement is reasonably exact in the aggregate, but precision cannot be guaranteed in an individual case. Reputable tests measure well what they are designed to measure, but that is not always what the public thinks they measure. As for the relevance of measures to purposes, it depends on the proper use of tests. For example, if a test unrelated to job requirements is used to screen potential employees, that test is being misused.

The issue of *fairness in testing* emerged as a prime concern of conference participants. Their doubts stemmed in part from popular protests that tests are culturally biased against low-income and minority groups and do not predict on-the-job or in-the-classroom performance accurately for members of those groups.

It would not be accurate to say that consensus was reached on these issues. There have been documented cases of test use for the purpose of discriminating against minorities, but they usually demonstrate that the tests were misused, not that they were planned instruments of oppression. The preponderance of available evidence indicates that, while minority group members and the poor do not score as high on the average as majority group members and the more affluent, test scores predict individual performance about equally well for all.

One of the few heated exchanges of the conference was related to the issue of cultural bias. One speaker asserted that tests are *value-neutral*. In the subsequent

discussion, a writer countered that tests reflect the values of the colleges doing the admitting or the industries doing the hiring. The values of these institutions in turn reflect the values of society in general. So the tests are *value-laden*, not value-neutral. But the speaker stood his ground: Tests measure the individual's knowledge or skill in a specified field at a particular time. Values are applied only by those who interpret and use test scores.

Subsequent informal conversations among participants indicated that the impasse was a matter of semantics. Tests are not value-free. They do reflect the values of society and its institutions. As one ETS staff member said, a test that is free of values, free of cultural influences, is not likely to measure anything of practical use. On the other hand, tests are value-neutral in that they are passive instruments that can be used for good or evil. The same test that bars someone from a school or a job might be used as a guide to remedial education or additional training.

Another concern about fairness in testing apparently sprang from apprehension that the increasing use of tests for occupational licensing and professional certification—and simply for hiring—would close more doors than it opened for minorities and the poor. A writer asked whether wider use of performance-based tests for persons entering various fields would not spark protests from those being tested. He said the tests were likely to be seen as barriers to economic opportunity just as college admission tests have been considered barriers to educational opportunity. To illustrate the point, he envisioned a young black man who graduates from a teacher-training university and suddenly discovers there is a new hurdle to jump before he can get his teaching certificate. In the past, a degree was all that was needed. Now a single test can block his way. Won't he feel that it's unfair?

Answers came from two ETS staff members. One said the nature of the tests will determine, to some extent, how much hostility they arouse. If a test is shrouded in secrecy, and failure is final, an adverse reaction is inevitable. But performance-based tests are being designed as diagnostic tools with feedback to the candidate so weaknesses can be corrected through further study or experience. The other staff member noted that, in any case, measurement agencies have an obligation to demonstrate that their tests are valid and to insure that they are "used in valid ways."

One journalist wanted to know who determines—or should determine—the criteria for measurement in a given case: Schools and colleges? The federal government? Occupational or professional practitioners? And what role does the measurement agency play? For instance, if questionable criteria are set forth, does ETS accept or challenge them? The answer was "any of the above" might contract with ETS or another organization to develop a testing, assessment, or evaluation program, and whoever initiates the program retains ultimate control of the criteria. But ETS certainly has—and uses—opportunities to influence its clients' decisions, and can refuse to develop a program it believes is invalid. A staff member added that occasionally ETS gets a chance to advise policy-makers generally, outside the framework of specific program development. The U. S. Department of Labor, for example, once asked ETS to examine how states and municipalities handle occupational licensing and suggest improvements.

That exchange gives a hint of two recurrent themes that threaded through the discussions. One was *the responsibility of measurement agencies* in general, and ETS in particular, *to various publics*. This concern first arose in an editor's question relating to the favorable ETS evaluation of *Sesame Street* for Children's Television Workshop (CTW). That study, the editor said, was criticized by some on the grounds that the wrong questions were asked, that CTW set the boundaries of the inquiry and ETS simply worked within them. Was that true? And if so, where does ETS' responsibility lie? Although he didn't say so specifically, he seemed to be asking whether ETS did or did not feel an overriding obligation to serve the interests of the viewing public.

The issue is complex, and so was the answer. ETS is obviously obligated to meet its client's requirements. Yet, beyond that, ETS does have, and recognize, a responsibility to the public interest. When the two are in conflict, ETS must try to resolve the differences. It must use reason and expertise to influence the client. In the case of *Sesame Street*, there was no such conflict. What CTW wanted was legitimate: an evaluation of *Sesame Street* in terms of whether it was meeting the goals set for it by CTW. ETS was not specifically asked to evaluate the goals to see "whether they were the right ones" or examine cost benefits. As performed the evaluation was legitimate and honest, and it certainly was not inimical to the public interest.

Echoes of this question, pinpointing a measurement agency's relative responsibility in the face of competing interests, sounded frequently and related to testing programs as well as research and evaluation. The majority of test-takers are students, most of them seeking admission to colleges, graduate schools, and professional schools. It was logical, therefore, that the related theme manifested at this conference was *the responsibility of measurement agencies to students.*

A key question, repeated in various ways, was asked by one editor: Is it ETS policy to provide students with full information about test scores and their proper interpretation and how colleges use them? The answer was yes—to the extent possible. Test-takers receive much information from ETS and the program sponsor about their own scores and how to interpret them. Details about how colleges and universities interpret and use test scores in making admissions decisions vary considerably from one institution to another. Thus each institution is responsible for informing students about its own procedures. In general both the sponsors and ETS want to protect the students' interests and provide them with all the facts possible. One staff member noted that much more information is available to students taking the College Board tests today than was available to the students of 10 years ago.

Two suggestions for change came from the participants. An editor asked if ETS might shift gears, in light of higher education's metamorphosis from a seller's to a buyer's market, and "test" colleges in order to provide a sort of shopper's guide for potential students. He acknowledged that ETS was not likely to find an organization to sponsor such an activity, but his question was serious. In answer, a staff member called attention to one such attempt, the College Locater Service

sponsored by the College Board and operated by ETS. Through this service, students submitted forms indicating the kind of institution they wanted to attend, and ETS matched their descriptions with data on file, then forwarded the information to them.

The other suggestion came from one of the respondents, a graduate student, who said testing agencies and students need to "come together" for their mutual benefit. Rather than a new service for students, she proposed a continuing interchange between agency personnel and students. It would help students understand testing and the uses of test scores, and it would help the agencies evaluate their programs. She advocated three types of activity: 1) Testing agencies should send exhibits and emissaries to the campuses; 2) they should hold conferences for representative students who would relay what they learned to their fellow students; and 3) students should be included in the agencies' decision-making and evaluation processes through appointment to trustee and advisory boards, and to research internships.

Another topic that figured prominently in the discussions was *the evaluation of educational programs and institutions*. One writer asked: if resources are too limited for comprehensive evaluation of a school system, is some better than none? The answer was yes, if the variables to be studied are carefully selected on the basis of their importance as well as the cost of measuring them. If cost alone is the determining factor, the school system may end up with mountains of trivial data. A subsequent question was: Would a "systems approach" to evaluation make usable any of the useless information now available in abundance in some schools? The response was yes, some of it might be rendered useful, but provision must be made to incorporate it into a continuing evaluation plan.

What can be done, an editor asked, about the resistance—from classroom teachers, for instance—to evaluation that seems to threaten the status quo? The answer: If evaluation is needed for sound educational reasons—and it is—then the resistance must be overcome. If the goal is to improve learning and good teachers feel threatened by the proposed process of evaluation, the process might be modified. Conversely, the good teachers, if they understand the purpose, may be persuaded to see the process differently.

Finally, conference participants paid considerable attention to *new kinds of tests to meet the needs of a changing society*. Trends and prospects have been examined in some of the papers in this volume, and there is no need to review them here. But a pertinent observation, expressed during a discussion session, does merit reporting. It was that, in the future, the techniques of measurement will be applied not only to aptitudes and achievement levels but to more subtle indicators of personal and intellectual functioning. It will become possible, in other words, to measure not only "what makes a good law student" but "what makes a good lawyer."

ROSTER OF PARTICIPANTS

Speakers

Samuel Ball, Research Psychologist, ETS

K. Patricia Cross, Senior Research Psychologist, ETS

Fred M. Hechinger, Editorial Board, *New York Times*

Winton H. Manning, Vice President, ETS

Samuel J. Messick, Vice President, ETS

Melvin M. Tumin, Professor of Anthropology and Sociology, Princeton University

William W. Turnbull, President, ETS

Respondents

Myron Corman, Department of Urban Education, University College, Rutgers University

Carl Crawford, Superintendent, Elementary District, Flemington-Raritan Regional Public Schools

Jerrold K. Footlick, Education Editor, *Newsweek*

Robert Glaser, Co-Director, Learning Research and Development Center, University of Pittsburgh

George Harris, Program Officer, Southern Regional Council, Inc.

Marsha Levy, Graduate Student, Yale University

Ruth Mancuso, Chairman, Committee on Evaluation, New Jersey State Board of Education

Carmen C. Scott, Assistant Program Director, ETS

Constance Shally, Content Specialist, Foreign Language Department, Hightstown High School

Session Chairmen

Richard S. Levine, Vice President, ETS

Robert J. Solomon, Executive Vice President, ETS

John Summerskill, Vice President, ETS

E. Belvin Williams, Vice President, ETS

Guests

Theodore Bleecker, Editor-in-Chief, *The New York Teacher*

Jenne K. Britell, Director, Information Services, ETS

Mary Cheuse, Assistant College Editor, *Glamour*

John A. Crowl, Managing Editor, *The Chronicle of Higher Education*

Walter A. Graves, Executive Editor, *Today's Education*

Anne S. Hopkins, Education Writer, *Time*

Gene I. Maeroff, Education Writer, *New York Times*

Loretta Hunt Marion, Editor, *Scholastic Teacher*

Carole Martin, Education Editor, Associated Press

Ted Miller, Associate Editor, *Changing Times*

John P. Smith, Co-Manager, Media Relations, ETS

Midge Turk, College Editor, *Glamour*

Melinda Upp, Assistant to the Director for Public Information, National Institute of Education

Jim Warren, Editor, *APA Monitor*

Heather Weiss, Editorial Board, *Harvard Educational Review*

Richard Yurko, News Editor, *The Dartmouth*

Conference Chairmen

Scarvia Anderson, Vice President and Director of the Atlanta Office, ETS

Jane Wirsig, Executive Director, Information Services and Publications, ETS

Conference Staff

Vivian E. Brandle, Administrative Assistant to Ms. Wirsig, ETS

Gail Carlin, Secretary to Ms. Britell, ETS

Loretta K. Eszlari, Secretary, Information Services, ETS